# Eco514—Game Theory
# Lecture 13: Repeated Games (2)

Marciano Siniscalchi

October 28, 1999

## Introduction

[Again, by and large, I will follow OR, Chap. 8, so I will keep these notes to a minimum.]

## Review of key definitions

Recall our three payoff aggregation criteria: discounting, i.e.

$$(u_i^t)_{t\geq 1} \succ_i (w_i^t)_{t\geq 1} \quad \Leftrightarrow \quad \sum_{t\geq 1} \delta^{t-1}(u_i^t - w_i^t) > 0$$

(also recall that the *payoff profile* corresponding to a stream $(u^t)$ is taken to be $(1-\delta)\sum_{t\geq 1}\delta^{t-1}u(a^t)$); limit of means:

$$(u_i^t)_{t\geq 1} \succ_i (w_i^t)_{t\geq 1} \quad \Leftrightarrow \quad \liminf_{t\to\infty} \sum_{t=1}^{T} \frac{u_i^t - w_i^t}{T} > 0;$$

and overtaking:

$$(u_i^t)_{t\geq 1} \succ_i (w_i^t)_{t\geq 1} \quad \Leftrightarrow \quad \liminf_{t\to\infty} \sum_{t=1}^{T} (u_i^t - w_i^t) > 0.$$

Also recall the definition of machines:

**Definition 1** *Fix a normal-form game $G$. A machine for Player $i \in N$ is a tuple $\mathcal{M}_i = (Q_i, q_i^0, f_i, \tau_i)$, where:*

*(i) $Q_i$ is a finite set (whose elements should be thought of as labels);*

*(ii) $q_i^0$ is the initial state of the machine;*

*(iii) $f_i : Q_i \to A_i$ is the action function: it specifies what Player $i$ does at each state; and*

*(iv) $\tau_i : Q_i \times A \to Q_i$ is the transition function: if action $a \in A$ is played and Player $i$'s machine state is $q_i \in Q_i$, then at the next stage Player $i$'s machine state will be $\tau_i(q_i, a)$.*

VISIT…

Note: OR actually allows for arbitrary state spaces. For most proofs, finite state spaces are enough. There is one exception, which I will note below.

**Definition 2** *Fix a normal-form game $G$. The* minmax *payoff for player $i$, $v_i$, is defined by* $v_i = \min_{a_{-i} \in A_{-i}} \max_{a_i \in A_i} u_i(a_i, a_{-i})$.

We are not going to worry about implementing mixtures of action profiles in this lecture, so I just remind you of the definition of enforceability:

**Definition 3** *Fix a normal-form game $G$. A payoff profile $u \in \mathbf{R}^N$ is* enforceable *(resp.* strictly enforceable*) if $u_i \geq v_i$ (resp. $u_i > v_i$) for all $i \in N$*

## Perfect Folk Theorems

The strategies in the proof of the Nash folk theorem (say, with discounting) call for indefinite punishment following a deviation from the equilibrium. This is fine in games such as the Prisoner's Dilemma, in which the minmax actions are actually an equilibrium (indeed, in PD, they are the only equilibrium!). That is, the threat of indefinite punishment is indeed credible.

| | $A$ | $D$ |
|---|---|---|
| $A$ | 2,3 | 1,6 |
| $D$ | 0,1 | 0,1 |

Figure 1: Indefinite Punishment is not credible

However, consider the game in Figure 1. The Row player can hold the Column player down to his minimum payoff by choosing D, but D is strictly dominated for her. Hence, while (2,3) can be enforced as a Nash equilibrium payoff profile in the infinitely repeated version of the game by assuming that players switch to (D,D) following a deviation, this would not work if we required subgame perfection, regardless of the payoff aggregation criterion.

### A warm-up exercise: Perfect Folk Theorems for Limit-of-Means

Thus, we must be smarter than that. A key intuition is that, after all, a deviator need not be punished forever, but only long enough to wipe out any gains from his deviation.

Formally, fix a game $G = (N, (A_i, u_i)_{i \in N})$ and let $M = \max_{i \in N, a \in A} u_i(a)$. Fix an outcome $a^*$ of $G$; clearly, a one-time deviation by Player $i$ cannot yield more than $M - u_i(a^*)$. Thus, if Player $i$ deviates, we need only punish him so as to wipe out this payoff differential; clearly, how long this will be depends on the payoff aggregation criterion.

Intuitively, limit-of-means should make things very simple, because, following a deviation, punishers face only a finite number of periods in which they must forego their equilibrium

payoff stream in order to punish Player $i$. Under limit-of-means aggregation, finite periods do not matter, so punishers are indifferent between punishing and not punishing.

There is only one subtlety: of course, the same argument applies to Player $i$: *no one-time deviation can be profitable for her*! So, what exactly are we trying to deter?

The answer is that, although no *finite* deviation from the infinite repetition of $a^*$ is profitable for Player $i$, the following strategy could be. Suppose that, in the putative equilibrium we wish to construct, a deviation is followed by $L$ rounds of punishment (you can think of $L' = 1$ if you wish). Thus, if Player $i$ deviates once, she gets an extra payoff of (at most) $M - u_i(a^*)$, but then loses $u_i(a^*) - v_i$ utils in each of the subsequent $L$ periods.

Now suppose $L$ is small, so that $M - u_i(a^*) > L[u_i(a^*) - v_i]$. For example, in the game of Figure 1, suppose that $i$ is the Column player, that $a^* = (A, A)$, and that $L = 1$. Then a deviation yields 3 utils, whereas one round of punishment costs 2 utils. Then, Player $i$ can adopt the following strategy: deviate, then play a best-response to $p_{-i}$ (in Figure 1, play $D$) for $L$ periods; then, as soon as play is supposed to return to $a^*$, deviate and best-respond to the minmax profile, and so on. This is a profitable deviation. [Observe that it is also a neat example of a game in which the one-deviation property holds!]

Thus, we must choose $L$ large enough so

$$\forall i \in N, \quad M - u_i(a^*) < L[u_i(a^*) - v_i]$$

In Figure 1, it is enough to choose $L' = 2$.

To complete the argument, we must specify what happens if more than one player deviates, or if somebody deviates from the punishment stage. As in the proof of the Nash Folk Theorem, multiple deviations lead to the lowest-index player being punished (they will not occur in equilibrium anyway, so players cannot hope to get away with deviating because somebody else will deviate, too).

Finally, if one or more punishers fail to punish, we simply disregard these further (unprofitable) deviations: again, in equilibrium they are not going to occur, so players cannot count on them to improve their predicament after a first deviation.

This proves:

**Proposition 0.1** *[OR, Proposition 146.2] Fix a game $G$. Any feasible, strictly enforceable payoff profile of $G$ is a subgame-perfect equilibrium payoff profile of the limit-of-means infinitely repeated version of $G$.*

### Machines

You will undoubtedly notice that describing these strategies verbally is awkward; doing so formally (as we have tried to do in the proof of the Nash Folk theorem) is even worse.

Machines can help. For instance, here is a set of machines that players can use to implement the strategies in the proof of Proposition 0.1: Player $i$ uses machine $\mathcal{M}_i = (Q, q^0, f_i, \tau)$ (where $Q, q^0, \tau$ are common to all players) defined as follows.

First, $Q = \{N, P(j,t)\}$. $N$ is the normal state, in which $a^*$ is played. $P(j,t)$ is the state in which $j$ is punished, and $t$ more rounds of punishment are required.

Second, $q^0 = N$. Third, $f_i(N) = a_i^*$, $f_i(P(j,t)) = p_{-j,i}$ if $j \neq i$, and $f_j(P(j,t)) = r_j(p_{-j})$. This should be obvious.

Finally, $\tau(\cdot,\cdot)$ is such that we remain in $N$ if nobody deviates, we switch from $N$ to $P(j,L)$ if $j$ is the lowest-index deviator, we always move from $P(j,t)$ to $P(j,t-1)$ if $t \neq 0$, and we always move from $P(j,0)$ back to $N$.

Easy!

**Discounting**

The strategy profile thus constructed is *not* a subgame-perfect equilibrium of the game in Figure 1 if payoffs are aggregated via discounting, for *any* discount factor. Suppose the Column player deviates: then the Row player is supposed to choose $D$ for 2 periods, and hence receive a flow payoff of 0 units. She will then receive 2 forever after. However, if she deviates to $A$, *nothing happens*: the Column player continues to choose $D$, so again the Row player can choose $A$. Obviously, she prefers (1,1,2,2,...) to (0,0,2,2,...)!

Thus, the key intuition is that we must somehow ensure that punishers are *willing* to punish. In principle, one could think of punishing punishers, but this may fail to work with discounting: essentially, second deviations might require longer punishment periods (because the burden of carrying out the first punishment lasts for $L$ periods, and not just one), third deviations might require even longer punishments, and so on. This is certainly the case in the game of Figure 1.

The alternative is to *reward* punishers. This leads to OR's Proposition 151.1.

I am only going to offer a few comments on the proof. The "conciliation" states $C(j)$ serve to reward punishers, in a way. However, this is subtle: we *never* go back to the Nash state $C(0)$. What happens is, if $j$ deviates and $i$ punishes him, then after punishment we move to $C(j)$, which $i$ prefers to $C(i)$. Otherwise, we go to $C(i)$ and stay there until somebody else deviates.[1] In my opinion, this is also a punishment of sorts, but of course you are welcome to differ.

Also: the remark about the first condition on $\delta$ being sufficient has to do with the fact that, after $L$ periods of punishment, we move to $C(j)$; since $u_i(a(j)) < u_i(a(0))$ by assumption, this is actually a further punishment (which, er, actually reinforces my interpretation, but never mind that). The point is that this punishment may not be enough, or may come

[1] Again, let me repeat this because I first got it wrong in class (but then you guys spotted me!): whenever we are in state $C(j)$, we remain there if somebody deviates, and move to the state $P(k,L)$ in which Player $k$'s punishment begins if $k$ deviates from $a(j)$. Thus, what supports $a(j)$ as a continuation equilibrium after a deviation by Player $j$ is the threat of further punishment; $a(j)$ need not be a Nash equilibrium *per se*.

too late, so we make sure that Player $j$ is hit really hard for $L$ periods before switching to the conciliation phase.

Finally, the second condition on $\delta$ has the following interpretation: by punishing Player $j$, Player $i$ potentially loses $M - u_i(p_{-j}, r_j(p_{-j}))$ for the $L$ periods following Player $j$'s deviation (be it a deviation from $a(0)$ or from whatever $j$ was supposed to play). On the other hand, after $L$ rounds of punishments, the game will switch to the $C(j)$ conciliation stage. Now, Player $i$ prefers to be in the $C(j)$ state than in the $C(i)$ state, by assumption, so if the discount factor is large enough, she will not deviate.

[The only subtle point is that she may actually deviate at any $t \in \{1, \ldots, L\}$ (where time is measured starting from the first stage *after* Player $j$'s deviation). If she deviates at $t$, she will actually be held down to $v_i$ starting from $t+1$ and until $t + L \geq L + 1$; the condition in the text is thus stronger than it needs be (it assumes a payoff of $M$ from $t+1$ to $L$, and a punishment payoff of $u_i(a(i))$ from $L+1$ to $t+L$).]